Imitation learning for clinical decision support in pediatric ECMO*

Fateme Golivand Darvishvand[1] **, Michael Skinner[1], Saurabh Mathur[2], Ameet Soni[3], Phillip Reeder[4], Kristian Kersting[2], Lakshmi Raman[4], and Sriraam Natarajan[1]

[1] University of Texas at Dallas
[2] Technische Universität Darmstadt
[3] Swarthmore College
[4] University of Texas Southwestern Medical Center

Abstract. Pediatric critical care is a dynamic, high-stakes process involving constant monitoring and adjustments in life-saving treatments. Modeling these interventions is crucial for effective decision support. In pediatric Extracorporeal Membrane Oxygenation (ECMO), where data are scarce and clinical dynamics are complex, we frame clinical decision-making as imitation learning from observational data, in which actions are not directly observed. We consider TabPFN, a recent transformer-based approach for tabular data, and traditional baselines including XGBoost and Multi-Layer Perceptrons on real-world pediatric ECMO data to learn the action models. We find that the TabPFN-based approach consistently outperforms these classical baselines, supporting its use as a strong clinician-behavior baseline for pediatric ECMO decision support.

Keywords: Pediatric ECMO · Imitation Learning · Clinical Decision Support · Tabular Transformers.

1 Introduction

Managing patients on life support requires clinicians to make continuous, high-stakes adjustments to machine settings in response to rapidly changing physiological signals. However, automated decision support remains limited, particularly in specialized domains like pediatric Extracorporeal Membrane Oxygenation (ECMO) [5], a method used to support critically ill children with cardiac or respiratory failure. Building decision support systems for this domain is complicated by the incomplete and difficult-to-formalize nature of pediatric-specific protocols, alongside the ethical and practical impossibility of experimentation on critically ill children.

We address this challenge of learning to manage pediatric ECMO by framing it as an imitation learning (IL) [4,7] task; here, an AI agent learns a policy that

---

[*] We acknowledge support from NIH award R01NS133142

[**] Corresponding author, `fateme.golivanddarvishvand@utdallas.edu`

maps observations to actions by mimicking expert behavior. Imitation learning from ECMO trajectories presents three key challenges, namely irregular clinical observation intervals, a concurrent, high-dimensional action space without explicit annotations, and severe data scarcity. We resolve these by aggregating trajectories into hourly windows, using expert knowledge to infer latent clinician actions, and employing TabPFN, a pretrained tabular transformer, as the base policy learner. We empirically evaluate this method on a small dataset of 78 pediatric ECMO trajectories, and find that the TabPFN-based learner outperforms traditional baselines such as a multilayer perceptron (MLP) and gradient-boosted trees, demonstrating the promise of tabular foundation models and IL for clinical decision support in this challenging setting[5].

2 Learning to act in Pediatric ECMO

We aim to learn policies to manage children on ECMO life support by mimicking expert behavior in historical trajectories. We define the task as follows:

Given: Observational pediatric ECMO patient trajectories.
Task: Learn a treatment policy that imitates clinician behavior.

We address this task using Imitation Learning (IL) [4]. Unlike reinforcement learning [9], which requires active interaction, offline IL uses expert demonstrations already available in the historical clinical records, making it more suitable for pediatric ICU settings. To formalize pediatric ECMO as IL, we first derive latent clinical actions from the patient trajectories using expert knowledge. We then decompose the imitation learning task into a factorized, hierarchical policy model, predicting interventions independently per action knob and separating the identification of a change from the determination of its direction. Finally, we instantiate the imitation learner using the TabPFN tabular foundation model and compare it with instantiations based on traditional machine learning models.

Each trajectory is represented as a sequence of state–action pairs, (s_t, a_t). The state s_t is constructed from four primary data streams: hemodynamics, laboratory test results, ECMO circuit variables, and ventilator settings. For each selected variable, we use both the mean hourly value and its delta relative to the preceding hour to capture short-term temporal trends. The action a_t is modeled as a multi-dimensional vector $(a_t^1, \ldots, a_t^K)$, allowing for the representation of simultaneous adjustments across K distinct "knobs". Since ECMO trajectories lack explicit intervention labels, we infer each knob-specific action k by monitoring its delta over a 1-hour window:

$$a_t^k = \begin{cases} \text{Increase}, & x_{t+1}^k - x_t^k > \delta_k, \\ \text{Decrease}, & x_{t+1}^k - x_t^k < -\delta_k, \\ \text{Same}, & \text{otherwise}, \end{cases}$$

---

[5] Code is available at https://github.com/fateme-gd/ImitationLearning_ECMO.git

where δ_k is a physician-defined threshold. This converts physiologic trajectories into supervised demonstrations while filtering small fluctuations that are unlikely to reflect meaningful intervention. This state–action definition allows us to make the first-order Markov assumption [8], approximating clinician behavior as the conditional distribution of a decision at time t given only the current state instead of the full history: $p(A_t \mid s_t)$.

To address the high-dimensional action space and the dominance of the "Same" action, we adopt a factorized, hierarchical policy structure. We train independent predictors for each knob, which allows the model to recommend concurrent interventions while avoiding the combinatorial explosion of a joint action space of 3^K possible composite actions. Each knob-specific head employs a two-stage process: first, a binary classifier identifies a change; second, conditional on a change, a classifier determines its direction.

Table 1: Actionable features (knobs) and physician-defined action thresholds. A discrete action is identified when the absolute change within a 60-minute window exceeds the specified threshold.

Knob	**Threshold**
Arterial PO_2	25 mmHg
Arterial PCO_2	5 mmHg
SpO_2	5 %
FiO_2	10 %

We instantiate this framework using the TabPFN 2.5 tabular transformer [3]. TabPFN encodes a broad tabular prior from its extensive pretraining on millions of synthetic datasets. This enables it to perform probabilistic inference in context, producing calibrated predictive probabilities in a single forward pass without requiring dataset-specific gradient-based training or hyperparameter tuning. We compare this instantiation with ones based on a 3-layer feedforward neural network (MLP) trained with gradient-based optimization [2] and a gradient-boosted model (XGBoost) consisting of 200 shallow trees [1].

3 Evaluation

We evaluate the generalization performance, calibration, and alignment with clinician decisions of the three imitation learners using a small pediatric cohort.

Dataset. We obtained 78 pediatric ECMO trajectories from Children's Medical Center, Dallas; these excluded cases with congenital heart disease due to the complexity of specialized management. The cohort had an average age of 4.66 years and an average trajectory length of 215 hours. The state vector consists of 48 features, encompassing 23 physician-selected clinical variables across hemodynamics, ECMO circuit configurations, ventilator settings, and laboratory results; their one-hour deltas; the ECMO type; and an active-support indicator. To account for pediatric developmental variance, heart rate and mean arterial pressure are age-normalized. Action labels are derived using the physiological thresholds defined in Table 1[6].

---

[6] A complete description of the state and action space is provided in https://github.com/fateme-gd/ImitationLearning_ECMO/blob/master/Supplementary.pdf

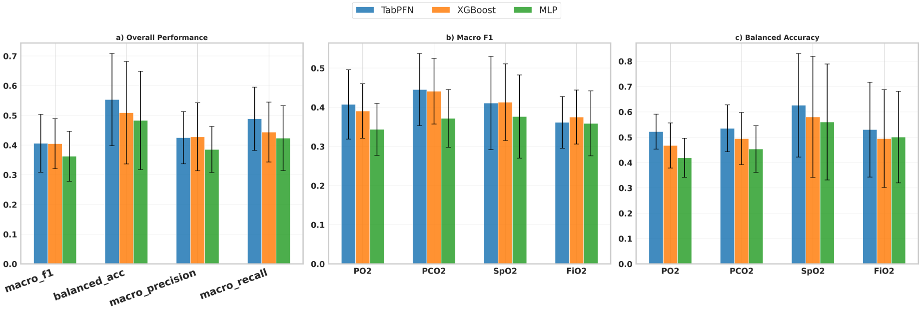

Fig. 2: Action selection performance of imitation learners based on TabPFN, XGBoost, and MLP. Metrics computed using leave-one-out cross-validation. (a) Aggregate performance. (b–c) Per-action-head performance. Values show mean ± standard deviation.

Metrics. Given the inherent class imbalance, we quantify the performance of the imitation learners using balanced accuracy and macro-F1 scores for each action head, alongside aggregate macro-precision and macro-recall. Additionally, we quantify model calibration using Expected Calibration Error (ECE) [6], where lower values indicate that the model's predicted confidence aligns more closely with its empirical accuracy. We compute these metrics using leave-one-out (LOO) cross-validation: in each fold, one patient trajectory is held out for testing, and all remaining patients are used for training. We apply the same LOO protocol and identical two-stage pipeline for all methods.

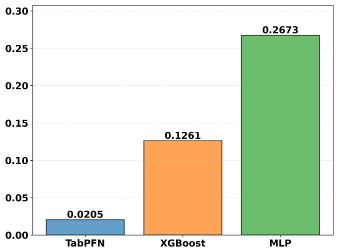

Fig. 1: Expected Calibration Errors of imitation learners based on TabPFN, XGBoost, and MLP.

Results. We now present our results, organized around three questions.

Q1: Do the learned policies generalize to unseen data? TabPFN achieves the highest mean balanced accuracy and macro-F1 across the majority of the actions (Fig. 2). The consistent improvement in balanced accuracy suggests that TabPFN is more resilient to the class imbalance inherent to this setting and generalizes more effectively to held-out patients.

Q2: Are the policies calibrated? Beyond raw accuracy, clinical decision support models must be well-calibrated to be trustworthy. TabPFN consistently achieves a lower ECE than the MLP and XGBoost baselines (Fig. 1). This indicates that the posterior probabilities produced by TabPFN's in-context inference closely match its empirical accuracy, providing reliable confidence estimates.

Q3: Where do the learned policies disagree with the clinicians? Beyond quantitative metrics, we investigate the specific physiological conditions under which the learned policies diverge from human expert decisions. To identify these regions of the state space, we extracted the most frequent features across states with the highest model-expert disagreement by training shallow decision trees on the residual errors of their predicted probabilities.

Oxygenation-related signals, specifically FiO_2 and SpO_2, consistently define the boundaries where model predictions deviate from clinician actions. However, the nature of these disagreements varies by model type: TabPFN disagreements are most often associated with FiO2, and SpO_2, whereas MLP disagreement more often involves PO_2 and lactate. This suggests that the models deviate from clinician actions in different regions of the state space and motivates future expert review of high-disagreement states.

Table 2: Most frequent features appearing in high disagreement states, across all knobs for each model.

Feature	**MLP**	**TabPFN**	**XGBoost**
FiO_2	2	5	5
SpO_2	1	5	3
pH	2	3	2
PO_2	3	2	1
PCO_2	2	3	1
Lactate	3	1	1

4 Conclusion

We presented an Imitation Learning framework to model clinical decision-making in pediatric ECMO. Our framework maps patient states to concurrent clinical interventions by inferring action labels from patient trajectories and employing a factorized, hierarchical policy architecture. Our empirical evaluation finds that TabPFN consistently outperforms traditional gradient-based baselines, suggesting that it can serve as an effective clinician-behavior baseline in challenging critical care settings. There are two key directions for future work. First, while we identify regions of model-clinician disagreements, future work should analyze whether they correlate with patient outcomes, such as risk of neurological injury. Second, while our policies accurately replicate clinician behavior, future work should explore policy improvement by incorporating expert-defined rewards.

References

1. Chen, T., Guestrin, C.: Xgboost: A scalable tree boosting system. In: KDD (2016)
2. Goodfellow, I., Bengio, Y., Courville, A., Bengio, Y.: Deep learning, vol. 1. MIT press Cambridge (2016)
3. Hollmann, N., Müller, S., Purucker, L., Krishnakumar, A., Körfer, M., Hoo, S.B., Schirrmeister, R.T., Hutter, F.: Accurate predictions on small data with a tabular foundation model. Nature **637**(8045), 319–326 (2025)
4. Hussein, A., Gaber, M.M., Elyan, E., Jayne, C.: Imitation learning: A survey of learning methods. ACM Computing Surveys (CSUR) **50**(2), 1–35 (2017)
5. Lin, J.C.: Extracorporeal membrane oxygenation for severe pediatric respiratory failure. Respiratory care **62**(6), 732–750 (2017)
6. Naeini, M.P., Cooper, G., Hauskrecht, M.: Obtaining well calibrated probabilities using bayesian binning. In: AAAI. vol. 29 (2015)
7. Natarajan, S., Joshi, S., Tadepalli, P., Kersting, K., Shavlik, J.: Imitation learning in relational domains: A functional-gradient boosting approach. In: IJCAI (2011)
8. Puterman, M.L.: Markov decision processes: discrete stochastic dynamic programming. John Wiley & Sons (2014)
9. Sutton, R.S., Barto, A.G.: Reinforcement Learning: An Introduction. MIT (2018)